Remote Sensing of Environment 206 (2018) 98–106

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Predicting the minimum height of forest fire smoke within the atmosphere using machine learning and data from the CALIPSO satellite

Jiayun Yao[a,*], Sean M. Raffuse[b], Michael Brauer[a], Grant J. Williamson[c], David M.J.S. Bowman[c], Fay H. Johnston[d], Sarah B. Henderson[a,e]

[a] School of Population and Public Health, University of British Columbia, Vancouver, Canada
[b] Air Quality Research Center, University of California, Davis, USA
[c] School of Biological Sciences, University of Tasmania, Hobart, Australia
[d] Menzies Institute for Medical Research, University of Tasmania, Hobart, Australia
[e] Environmental Health Services, British Columbia Centre for Disease Control, Vancouver, Canada

ARTICLE INFO

Keywords:
Forest fire smoke
CALIPSO
Vertical profile
Machine learning
Statistical model
Population exposure

ABSTRACT

Forest fire smoke is a growing public health concern as more intense and frequent fires are expected under climate change. Remote sensing is a promising tool for exposure assessment, but its utility for health studies is limited because most products measure pollutants in the total column of the atmosphere, and not the surface concentrations most relevant to population health. Information about the vertical distribution of smoke is vital for addressing this limitation. The CALIPSO satellite can provide such information but it cannot cover all smoke events due to its narrow ground track. In this study, we developed a random forests model to predict the minimum height of the smoke layer observed by CALIPSO at high temporal and spatial resolution, using information about fire activity in the vicinity, geographic location, and meteorological conditions. These pieces of information are typically available in near-real-time, ensuring that the resulting model can be easily operationalized. A total of 15,617 CALIPSO data blocks were identified as impacted by smoke within the province of British Columbia, Canada from 2006 to 2015, and 52.1% had smoke within the boundary layer, where the population might be exposed. The final model explained 82.1% of the variance in the observations with a root mean squared error of 560 m. The most important variables in the model were wind patterns, the month of smoke observation, and fire intensity within 500 km. Predictions from this model can be 1) directly applied to smoke detection from the existing remote sensing products to provide another dimension of information; 2) incorporated into statistical smoke models with inputs from remote sensing products; or 3) used to inform estimates of vertical dispersion in deterministic smoke models. These potential applications are expected to improve the assessment of ground-level population exposure to forest fire smoke.

1. Introduction

Smoke emitted from forest fires is a major contributor to poor air quality in many parts of the world. Fire smoke is a complex mixture of pollutants, among which particulate matter (PM) is the most consistently elevated and commonly studied (Durán et al., 2014; Naeher et al., 2007). Given that climate change favors more intense and frequent fires, smoke emissions are projected to increase over the coming decades (Spracklen et al., 2009). Under these conditions, exposure to forest fire smoke is becoming a pressing public health concern with short- and long-term air quality impacts affecting an increasing number of people (Liu et al., 2016). Forest fire smoke exposure has been consistently associated with a wide range of acute respiratory health endpoints, from increased medication dispensations to hospital admissions and mortality. Evidence for cardiovascular endpoints is also emerging (Liu et al., 2015; Reid et al., 2016a). However, better exposure assessment is needed to improve epidemiologic studies of these effects (Atkinson et al., 2014; Pope III and Dockery, 2006).

It is challenging to assess population exposure to forest fire smoke both retrospectively and in near-real-time because smoke is dynamic and variable in both space and time. Good exposure assessment requires tools with high spatial and temporal resolution, as well as sufficient spatial coverage. Conventional ground-based air quality monitoring can be limited with respect to these spatial considerations, but remote sensing technology has provided promising alternatives. Remotely sensed fire and/or smoke products have been used to directly or

* Corresponding author at: School of Population and Public Health, University of British Columbia, 2206 East Mall, Vancouver V6T 1Z3, Canada.
E-mail address: Jiayun.Yao@bccdc.ca (J. Yao).

https://doi.org/10.1016/j.rse.2017.12.027
Received 16 March 2017; Received in revised form 5 December 2017; Accepted 18 December 2017
Available online 22 December 2017
0034-4257/ © 2017 Elsevier Inc. All rights reserved.

indirectly assess smoke exposure in many recent studies (Faustini et al., 2015; Henderson et al., 2011; Kollanus et al., 2017; Reid et al., 2016b; Yao et al., 2016). However, most remote sensing products provide information integrated from the top to the bottom of the atmosphere, rather than information specifically at the surface where populations are exposed. Better information about the vertical distribution of smoke in the atmosphere at adequate temporal resolution would improve the utility of remote sensing products for exposure assessment in epidemiologic and public health surveillance applications.

There are two remote sensing platforms that can provide information about the vertical distribution of atmospheric aerosols. The Multi-angle Imaging Spectro-Radiometer (MISR) aboard the Terra satellite (Kahn et al., 2007) can measure the altitude of the layer of maximum contrast, which can be the land surface, cloud top, or smoke plume top. However, MISR does not always provide the complete vertical profile of the smoke plume, especially for thin smoke or smoke far from the source (Kahn et al., 2008). On the other hand, products derived from the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite can provide both aerosol feature classifications (e.g. smoke, dust, etc.) and a complete picture of the vertical profile. However, data from CALIPSO are limited by its extremely narrow swath of observations. Given that CALIPSO passes over most locations only once every 16 days, it is unable to measure the vertical profile of most smoke plumes.

To achieve more spatially and temporally resolved information on the vertical distribution of smoke we need to turn to models rather than measurements. One approach is deterministic modelling, which relies on equations of the physical process of fire and smoke, as well as atmospheric transport (Paugam et al., 2016). In these models, the initial maximum smoke plume rise over the fire (injection height) and the subsequent vertical distribution of smoke (downwind height) are modelled separately, with the former as an important input for the latter. Such models are currently used for operational smoke forecasting systems, including the BlueSky (Larkin et al., 2010; Sakiyama, 2013) and FireWork (Pavlovic et al., 2016) frameworks in Canada. So far, most studies have focused on using CALIPSO and MISR data to evaluate and improve the injection height estimate, which describes the highest altitude of the smoke emission near the fire locations (Raffuse et al., 2012; Sessions et al., 2011; Sofiev et al., 2012; Val Martin et al., 2010). However, the more important estimate for public health is the downwind height, which describes smoke dispersion to the surface where human populations can be exposed.

Another approach is statistical modelling, which uses empirical data to describe the statistical relationships between relevant variables without considering the intermediate physical processes. We previously developed the near-real-time Optimized Statistical Smoke Exposure Model (OSSEM) to operationally estimate daily PM concentrations across the Canadian province of British Columbia (Yao et al., 2016; Yao and Henderson, 2014) using remote sensing data. However, the initial implementation of OSSEM did not account for the vertical profile of smoke in the atmosphere. This limitation may prevent improvement in the model performance, especially when we try to expand the model to finer temporal resolution. To address this limitation, we must develop a complementary near-real-time model that can operationally estimate the minimum height of smoke in the column. To the best of our knowledge there are no such models for injection height or downwind height, although some studies have recognized their feasibility and have identified potentially predictive variables related to fire intensity and meteorological conditions (Freitas et al., 2006; Freitas et al., 2007; Kahn et al., 2007; Mazzoni et al., 2007; Peterson et al., 2014; Val Martin et al., 2012). The process of building such models is further facilitated by the rapidly developing field of machine learning, where new methods can be used to model relationships that are far more sophisticated and complex than those that can be captured by conventional regression. The simplicity of statistical modelling with easily accessible data can provide timely estimates in the operational setting, especially in the situation where computational power or expertise to develop and maintain a complex deterministic model is lacking. However, most machine learning approaches come with a major limitation that the relationships between variables and their physical meanings are very difficult to properly assess.

In this study, we applied random forests to predict the minimum height of the smoke layer observed by CALIPSO using information about fire activity in the vicinity, geographic location, and meteorological conditions. These pieces of information are typically available in near-real-time, ensuring that the resulting model can be operationally applied. The predictions can be 1) directly applied to smoke detections from the existing remote sensing products; 2) incorporated into statistical models with inputs from remote sensing products; or 3) used to inform estimates of injection height and downwind dispersion in deterministic models. These potential applications are expected to improve the assessment of ground-level population exposure to forest fire smoke for both epidemiologic research and public health surveillance.

2. Methods

2.1. Study setting

The study area is the province of British Columbia (BC), on the west coast of Canada. With approximately 70% of its land area covered by forests, BC is prone to forest fires (B.C. Ministry of Forests, 2010). In recent years a severe pine beetle infestation has left approximately 18 million hectares even more susceptible, and several of the worst fire seasons on record have occurred in the past decade (Natural Resources Canada, 2016). Smoke from these fires has caused episodes of very poor air quality in many parts of the province during the fire season, which we have defined as April through September based on historical data. This study covers all forest fire seasons between 2006 and 2015, when data were available from all sources.

2.2. The CALIPSO data and the response variable

The CALIPSO satellite was launched in April 2006, as part of the US National Aeronautics and Space Administration (NASA) Afternoon Constellation (A-train), in a polar orbit with a 16-day repeat cycle (Fig. 1A). The primary instrument carried by CALIPSO is the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP). The CALIOP instrument generates laser pulses at two wavelengths (532 nm and 1064 nm) every 333 metres (m) along the satellite ground track, and then collects the returned backscatter signals. Analysis of these signals allows for discrimination between cloud and aerosol, and provides insight into the vertical distribution, shape, size, and type of detected aerosols (Winker et al., 2006).

We obtained the CALIPSO Level 2 lidar vertical feature mask data (Version 3) from the CALIPSO Search and Subsetting Web Application for all overpasses in BC during the study period. These data were stored in a block for each 5-kilometre (km) segment of the footprint along the swath (Fig. 1A and B), with data cells at different vertical and horizontal resolutions depending on the altitude (Fig. 1B and C). The vertical resolution is 30 m for all altitudes below 8.2 km, which are the most relevant to human activity and exposure. A detailed description of the algorithm used to classify features of the CALIPSO data can be found elsewhere (Liu et al., 2005; Liu et al., 2009; Omar et al., 2009; Vaughan et al., 2005; Vaughan et al., 2009). In brief, each data cell is classified as *cloud*, *aerosol*, or *other*, with a quality assessment (QA) of *low*, *medium*, or *high*. The distinction between *cloud* and *aerosol* is primarily based on the scattering strength and the spectral dependence of backscattering. If a data cell is classified as *aerosol*, its type is then further classified into *clean marine*, *dust*, *polluted continental*, *clean continental*, *polluted dust*, *smoke*, or *other* (Fig. 1), with a binary QA of *confident* or *not confident*. Aerosol subtype classifications are primarily

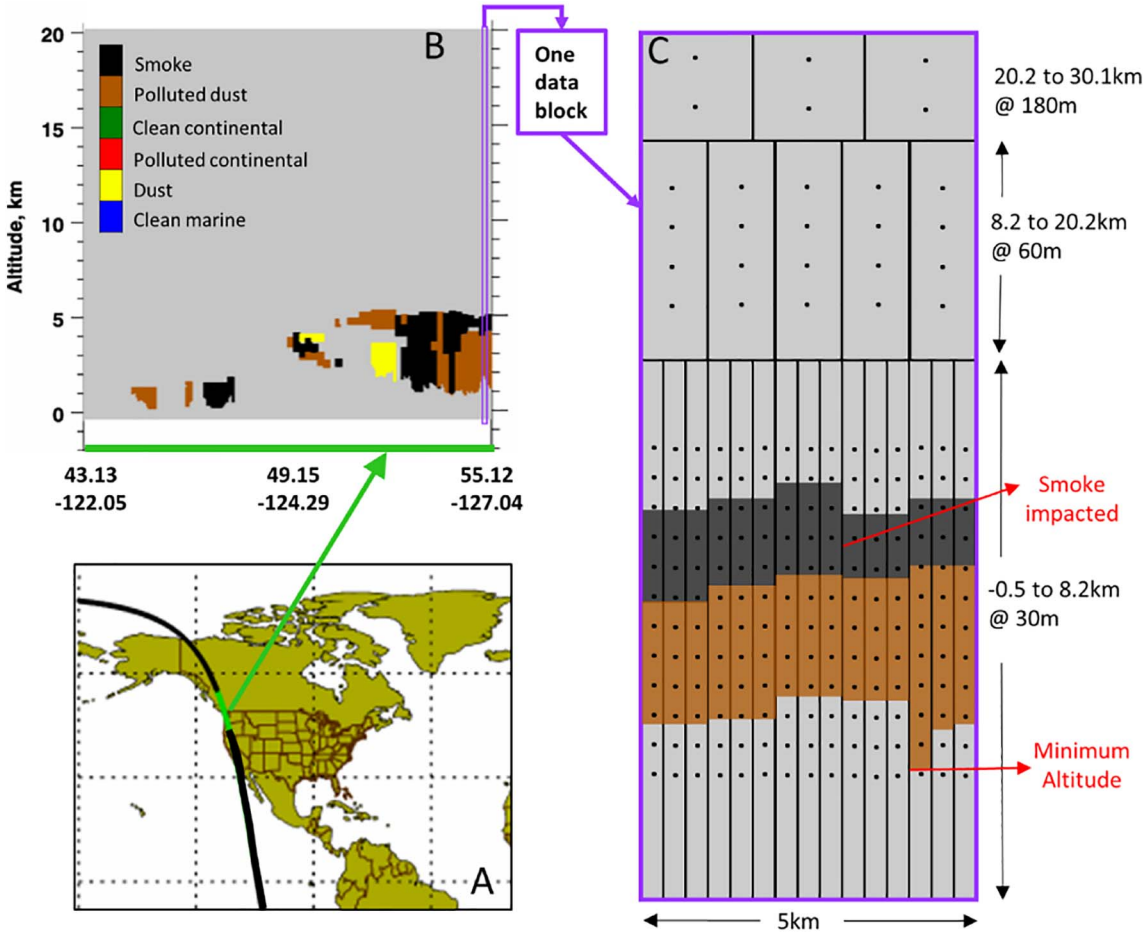

Fig. 1. Illustration of an example of the footprint (A), the aerosol classification and vertical profile (B), and a data block (C) of the CALIPSO data. Panel C also shows the definition of smoke impacted block and the minimum altitude of aerosol in the block.
Adapted from the CALIPSO product description (US NASA, 2016).

based on indicators such as depolarization, geophysical location, and backscatter intensity.

We defined our *Minimum Height* response variable as the minimum height of detected smoke above the land surface. To construct this variable, the first step was to identify all the CALIPSO data blocks that were impacted by smoke. A data block was considered to be smoke-impacted when the following two conditions were met: 1) any data cell within the data block was classified as *aerosol* with a *high* QA value, and as subtype *smoke* with a *confident* QA value; and 2) there were fire hotspots observed within 500 km of the data block within a 24-hour window of the smoke classification, spanning 12 h on either side of the overpass. The fire hotspots were detected using data from the Moderate Resolution Imaging Spectroradiometer (MODIS), as described in the following section. The second step was to find the minimum altitude of the smoke layer. In each smoke-impacted data block, the *Minimum Altitude* variable was defined as the altitude of the lowest cell classified as *aerosol* above sea level, regardless of its subtype classification (Fig. 1C). This approach was used because the subtype classification algorithm requires aerosol to be lofted before it can be classified as *smoke* (Liu et al., 2005). As a result, any smoke in the surface layer would have been classified as another type of aerosol. Given the limited sources of summertime aerosol pollution in most of the province and the fire hotspots requirement, we assumed that any aerosol underlying smoke in the 5-km data block was also smoke or a mixture of smoke and other aerosols. Finally, the *Minimum Height* variable was created by subtracting the elevation of the land surface at the location of the smoke detection from the *Minimum Altitude* variable to reflect the distance of the smoke layer from the land surface instead of the sea level.

2.3. Other data and the potentially predictive variables

Because the horizontal resolution of the CALIPSO data blocks was 5 km, all of the data for deriving the potentially predictive variables were first rasterized into a base grid of 5 km × 5 km over the province, and then the value of the cell in which the specific smoke detection fell was extracted for analysis. We selected variables that were expected to be associated with the outcome, commonly available near real-time, and easy to understand even for users outside the field of atmospheric science or meteorology. These variables were classified into four categories, as described below.

2.3.1. Fire activity

Information about active fires was retrieved from the MODIS active fire product (MCD14ML, Collection 6), distributed by the file transfer protocol (FTP) server maintained by the University of Maryland (LANCE FIRMS, 2016). The MODIS instruments measure thermal infrared brightness temperature of the land surface, and active fires are identified where thermal abnormalities are detected in 1 km pixels (Giglio et al., 2003). There are MODIS instruments aboard the Terra and Aqua satellites, each of which overpasses the province twice daily. The MODIS fire product provides the latitude and longitude of the fire pixels, the date and time of the fire detection, and its fire radiative power (FRP). The FRP is a quantitative measure of radiant heat output commonly used to approximate fire intensity, which is proportional to its combustion rate and smoke emissions (Ichoku and Kaufman, 2005; Wooster et al., 2003; Wooster et al., 2005). Fire hotspots were extracted from the database and rasterized to the 5 km base grid if they were (1)

Table 1
Description of the response and potentially predictive variables, as well as summary statistics.

Continuous variables		
Variables	Median [quartiles]	Description
Minimum Height (m)[a]	293 [59, 1416]	Response variable. Minimum height of aerosol above land surface in smoke-impacted CALIPSO data block
PBLH (m)	458 [138, 1310]	Planetary boundary layer height above land surface
Elevation (m)[a]	1016 [749, 1293]	Elevation of the land surface above sea level
Latitude (°)[a]	54.21 [50.795, 57.13]	Latitude of the smoke-impacted data block.
Longitude (°)[a]	−123.4 [−126.1, −121.1]	Longitude of the smoke-impacted data block
FRP Within 500km (MW)[a]	161 [33, 769]	Summed FRP for all fire cells within 500 km
Closest FRP (MW)[a]	15 [7, 38]	Summed FRP for the closest fire cell cluster
IDW FRP (MW)[a]	29 [13, 77]	The inverse distance weighted averaged FRP for fire cells within 500 km
Minimum Distance (km)	214 [115, 340]	Distance to the closest fire cell
Wind-Weighted Closest FRP (MW)[a]	13 [7, 35]	Summed FRP for the fire cell cluster with the shortest wind-direction cost-weighted distance
Wind-Weighted Minimum Distance (cost unit)	51 [28, 84]	The accumulative cost distance from the *closet FRP weighted* to smoke
E50m (m/s)[a]	1.37 [−0.72, 2.93]	Eastward wind at 50 m above surface
N50m (m/s)[a]	0.22 [−1.20, 1.70]	Northward wind at 50 m above surface
E850hPa (m/s)	1.84 [−0.63, 4.10]	Eastward wind at 850 hPa pressure level
N850hPa (m/s)[a]	0.15 [−2.00, 2.37]	Northward wind at 850 hPa pressure level
E500hPa (m/s)	5.65 [1.35, 10.40]	Eastward wind at 500 hPa pressure level
N500hPa (m/s)[a]	0.52 [−4.49, 6.53]	Northward wind at 500 hPa pressure level
E250hPa (m/s)[a]	9.74 [1.73, 19.26]	Eastward wind at 250 hPa pressure level
N250hPa (m/s)	−0.94 [−10.48, 8.49]	Northward wind at 250 hPa pressure level

Categorical variables			
Variables		N (%)	Description
Direction[b]	[0, 90]	4160 (26.6)	Direction as degrees bearing to the closest fire cell
	[90, 180]	5129 (32.9)	
	[180, 270]	3263 (20.9)	
	[270, 360]	3065 (19.6)	
Daytime	Day	3592 (23.0)	Binary variable indicating whether smoke was detected in daytime or nighttime overpass
	Night	12,025 (77.0)	
Month[a]	April	1392 (8.9)	Categorical variable indicating the month of the smoke detection
	May	1867 (12.0)	
	June	1405 (9.0)	
	July	3892 (24.9)	
	August	4746 (30.4)	
	September	2315 (14.8)	

[a] Variables included in the final model.
[b] Direction was calculated as degrees bearing (0 to 360) and used as a continuous variable in the model. It is categorized here for a more sensible descriptive summary.

observed 12 h before or after each CALIPSO smoke detection and (2) classified as *presumed vegetation fire* in the fire product. Both the scan angle and the bowtie effect can erroneously increase FRP near the edge of the satellite image swath (Freeborn et al., 2011; Freeborn et al., 2014). To account for this, we first calculated the scan angle and pixel size for each fire detection using the number of sampled pixels provided in the fire product (Giglio, 2015). The summed FRP assigned to each 5 km grid was then adjusted for the scan angle and pixel areas using previously described methods (Kaiser et al., 2012). The detailed methods for these adjustments are presented in the Supplementary materials Section 1.

Five potentially predictive variables were derived from the rasterized FRP data for each smoke detection (Table 1). The *FRP Within 500km* was calculated as summed FRP values for all fire cells within 500 km of the smoke detection. This variable was also used to evaluate the second criterion for defining smoke-impacted data blocks described in the previous section, where only observations with *FRP Within 500km* larger than zero were included in the analysis. The *Closest FRP Cluster* was the summed FRP value for the fire cell cluster closest to the smoke detection. A fire cell cluster was defined as a patch of neighboring fire cells that shared any boundary using the Queen's case criterion (Lloyd, 2010). The *IDW FRP* applied the inverse distance weighted (IDW) average to all fire cells within 500 km of the smoke detection, where each fire cell was given a weight as the inverse of its distance to the smoke observation (Shepard, 1968). This variable was created to reflect the fact that closer fires may have a larger impact on local smoke than more distant fires. The *Minimum Distance* and *Direction* variables indicate the distance and the direction in degrees bearing from the smoke detection to the nearest fire cell, respectively.

2.3.2. Meteorology

Meteorology plays an important role in smoke emissions and dispersion. As we were interested in developing an empirical model that could be operationalized in near real-time with readily available data, we did not attempt to retrospectively simulate the complex processes of fire behavior and pollutant transport as has been done in deterministic models. Instead we used a limited set of meteorological parameters accessible in near real-time and applicable in our machine learning approach.

The planetary boundary layer (PBL) is the lowest part of the atmosphere that is directly influenced by the surface of the earth, where turbulent air flow and vertical mixing are usually strong. The height of the planetary boundary layer (PBLH) above the surface varies in space and time, with a strong diurnal cycle over land (Stull, 2012). We retrieved hourly PBLH estimates from the two-dimensional surface turbulent flux diagnostics produced by the NASA Modern Era Retrospective-analysis for Research and Applications (MERRA) program (Rienecker et al., 2011) at a spatial resolution of 2/3-degree longitude by 1/2-degree latitude. The raw data were averaged to the 5 km base grid to generate a *PBLH* variable value in each cell.

A similar approach was used for the eastward and northward wind components at 50 m above the surface (*E50m* and *N50m*), respectively, to represent the wind patterns close to the surface. We also created variables from the eastward and northwards components of winds at three different air pressure levels: 850 hPa (*E850hPa*, *N850hPa*); 500 hPa (*E500hPa*, *N500hPa*); and 250 hPa (*E250hPa*, *N250hPa*). These variables can represent wind patterns at the lower, middle and upper troposphere, respectively. All wind data were retrieved from the two-dimensional atmospheric single-level diagnostics in the MERRA product, also at a spatial resolution of 2/3-degree longitude by 1/2-degree latitude.

2.3.3. Fire-meteorology combined variable

The *Wind-Weighted Minimum Distance* and *Wind-Weighted Closest FRP Cluster* variables were created to link fire, smoke, and meteorological conditions together. Because the movement of smoke can be facilitated or impeded by the wind field, these two variables calculated (1) the minimum distance to the closest fire cluster and (2) the FRP of that fire cluster, accounting for whether the relative locations of fire and smoke were aligned with the wind direction. The calculations included the following steps: (1) for each smoke observation, the direction from each fire cluster detected 12 h before and/or after was calculated in degrees bearing; (2) trigonometry was used to construct a 5 km × 5 km wind direction raster (in degrees bearing) using the two 850 hPa Cartesian wind components (*E850hPa* and *N850hPa*) at the time of the smoke observation; (3) a cost surface was created based on the difference between the fire-to-smoke direction and the wind direction in each grid cell, where larger distances had a higher cost (i.e. differences of 0–30, 30–60, 60–90, 90–120, 120–150, 150–180° were assigned cost values of 1 through 6, respectively); and (4) the least-cost distance between the smoke observation and each fire was calculated using the *costDistance* function in the *gdistance* package in R (van Etten, 2012). The ultimate result is that FRP clusters upwind of the smoke observation are closer in weighted distance than those that are downwind but at the same geographic distance. The FRP of the fire cluster with the smallest cost-weighted distance was assigned as *Wind-Weighted Closest FRP Cluster*, and the value of the cost distance was assigned as the *Wind-Weighted Minimum Distance* for the smoke observation.

Daytime and *Month* variables were created to capture further diurnal and seasonal patterns in fire activity and meteorology not covered by existing variables. Because CALIPSO only passes over BC from 01:00–03:00 (night) or 12:00–13:00 (day) local time, the *Daytime* variable cannot capture the full range of nighttime or daytime conditions under which smoke and fires occur.

2.3.4. Geographic location

The elevation of the land surface was obtained from the GTOPO30 product, a global digital elevation model with a horizontal grid spacing of 30 arc sec (approximately 1 km at the equator), developed by the US Geological Survey Earth Resources Observation and Science (EROS) Center (EROS Data Center, 1996). The raw data were assigned to the 5 km base grid as the average to generate values for the variable *Elevation* in each cell. The latitudes and longitudes of the smoke-impacted CALIPSO data blocks were used to create the *Latitude* and *Longitude* variables, which could reflect other important geospatial information, such as terrain features and land use.

Because our model was developed with the intention of future operational use in near-real-time, we ensured that all variables we included could be derived in near-real-time. Although we used meteorological variables from retrospective MERRA project for the model described here, comparable variables could easily be derived from operational weather forecasting systems. Similarly, the FRP data used in our model were derived from a product with retrospective quality assurance procedures, typically available months after the actual retrieval date, but the Fire Information for Resource Management System (FIRMS) maintained by NASA also provides a near-real-time FRP product. To operationalize this model, we could simply replace the retrospective data sources with their near-real-time substitutes, although further evaluation would be needed to assess model performance with data of different quality.

2.4. Statistical modelling

All data cleaning and analysis was conducted in the R Statistical Computing Environment (R Core Team, Vienna, Austria). The data were fitted with the random forests algorithm, an ensemble of regression trees, each of which was constructed with a random subset of observations and a random subset of predictive variables. This machine learning approach can provide accurate prediction while being robust against overfitting, accommodating non-linear relationships between the dependent and independent variables, and accounting for complex interactions between the independent variables (Breiman, 2001). We fitted a random forests model with 500 regression trees and a subset of four predictive variables sampled for each tree, using the *randomForest* package in R (Liaw and Wiener, 2002). Each tree was trained with a selected subset of data, and predictions were made with the subset of data that were not sampled for training, otherwise known as the "out-of-bag" data.

Because each data point fell into the out-of-bag subset multiple times out of the 500 trees, it had multiple predicted values for the *Minimum Height* response variable. The average of these values was reported as the final out-of-bag prediction for that data point, and the final out-of-bag predictions for all data points were used to calculate the root mean square error (RMSE) and the R-squared (percentage of variance explained by the model) against the observations. These values were used to evaluate the model performance.

The importance of each variable was calculated using a permutation test on the out-of-bag predictions (Breiman, 2001). If a predictive variable is important, a random permutation of the variable will considerably increase the model prediction errors when compared with a model that uses the real values of the variable. Thus, the increase in the RMSE between the permuted and actual model predictions, scaled by the standard deviation of the difference, was used to rank the importance of the variables.

A backward selection process was applied to select the final model. All variables were included in the initial model, then a reduced model excluding the least important variable would be fitted. If the RMSE and R-squared values for the reduced model were not significantly different from those for the initial model, the variable would be permanently excluded. The process was then repeated with the least important variable in the new model, which was the reduced model from the previous step. The process stopped and the final model was determined when the removal of the least important variable resulted in any reduction of R-squared values larger than 1% of the highest R-squared achieved in the previous models. This process was aimed to obtain the most predictive model with as few variables as possible, which will be useful for operationalizing the model in near real-time.

The out-of-bag predictions from the final model were also used to evaluate its accuracy, sensitivity (true positive rate), and specificity (true negative rate) in predicting whether *Minimum Height* was below or above thresholds of 100, 300, 500 and 1000 m. These thresholds cover the range of possible human exposure. Accuracy, sensitivity and specificity were calculated as the percentage of correct predictions among all observations, among observations below the threshold, and among observations above the threshold, respectively.

3. Results

A total of 15,617 CALIPSO data blocks met our criteria for classification as smoke-impacted and had complete information for all the potentially predictive variables. They were identified on 887 of the 1761 dates in the study period, and more than half (55.3%) occurred in

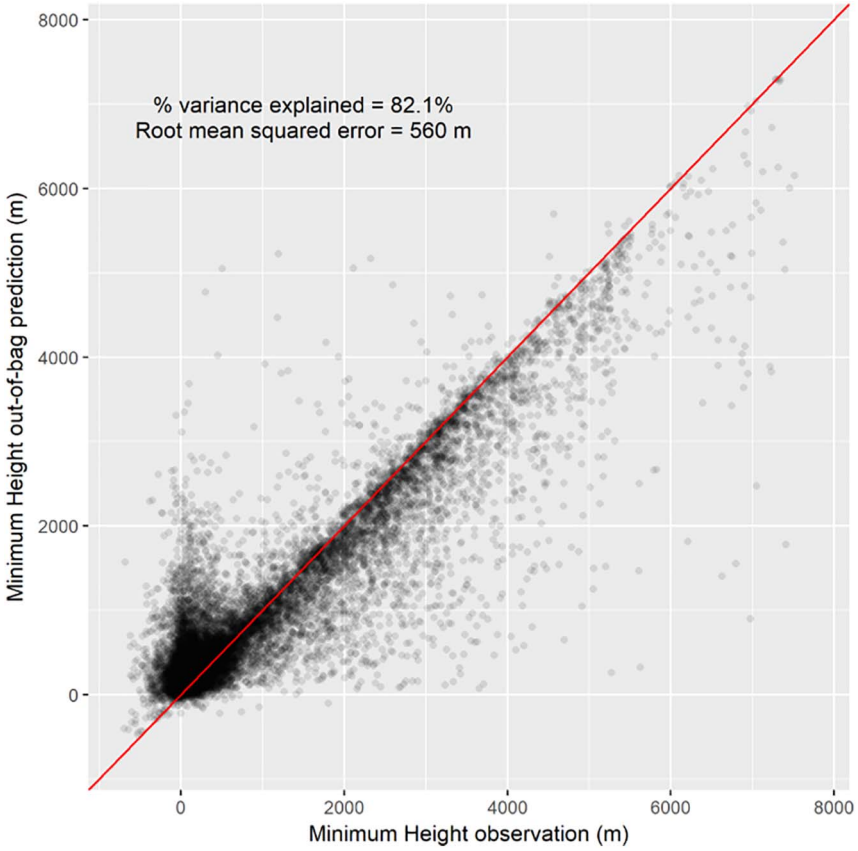

Fig. 2. Scatterplot comparing *Minimum Height* observations and out-of-bag predictions from the random forests model. The red line represents the 1:1 perfect match. When *Minimum Height* observations were below 500 m, there were generally more over-predictions (data points above the red line), while under-predictions (data points below the red line) dominate in other cases. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

July and August (Table 1), which is the peak forest fire season in the province. Furthermore, more than half (52.7%) had *Minimum Height* values below the PBLH at the corresponding time and location, meaning that smoke was more likely to be mixed in the boundary layer and to be affecting surface air quality.

The final random forests model after the backward selection process included 13 of the 21 variables (Table 1). The final model explained 82.1% of the variance in the out-of-bag *Minimum Height* observations, and the RMSE between the out-of-bag predictions and the observations was 560 m (Fig. 2), which was large compared with the vertical resolution of the CALIPSO data (30 m at minimum). There were generally more over-predictions when the observed *Minimum Height* values were below 500 m. A descriptive analysis of the data in the cluster of over-predictions with observed *Minimum Height* below 500 m found that they tended to be at lower elevation, to be farther from fires, and to have lower PBLH values when compared with the other observations (Supplementary materials Table S1). On the other hand, under-predictions were more uniform across the range of observed values.

In addition, the model predicted whether *Minimum Height* was below the 100 m, 300 m and 500 m thresholds with low to moderate sensitivity (0.31–0.80) and high specificity (0.90–0.96). This means that when the observed *Minimum Height* was below the respective thresholds, the model predicted correctly 31% to 80% of the time. Meanwhile, when the observed *Minimum Height* was above the thresholds, the model predicted correctly 90 to 96% of the time. In general, the accuracy and specificity increased considerably as the threshold increased, while the sensitivity decreased slightly (Table 2). In other words, the model became more accurate at predicating smoke below the threshold as the threshold height increased, and the model was quite accurate at predicting smoke above the threshold for all threshold values. This indicates that the model is especially useful for ruling out smoke that is not relevant to surface exposures.

The most important variable in the model was the speed of eastward wind at 50 m above the surface. There was a 97% increase in the normalized mean squared error when predictions were made using the randomly scrambled version of this variable compared with predictions using the actual values. The three most important variables in the model were all related to wind pattern, followed by variables related to fire activity, including *Month* and *FRP Within 500km* (Fig. 3).

Table 2
Model performance for predictions of *Minimum Height* below different threshold values.

	Minimum Height <			
	100 m	300 m	500 m	1 km
Accuracy	0.75	0.78	0.84	0.91
Sensitivity	0.31	0.65	0.80	0.94
Specificity	0.96	0.91	0.90	0.85

Accuracy = Percent of correction predictions among all observations.
Sensitivity = Percent of correct predictions among observations over the threshold.
Specificity = Percent of correct predictions among observations below the threshold.

4. Discussion

A machine learning approach was applied to predict the minimum height of smoke in the atmosphere using variables that reflected fire activity, geographic location, and meteorology, and the combination of fire activity and meteorology. The model explained 82% of the variability in the observations. The comparison of predictions and observations with binary categories for smoke under 100, 300, 500, and 1000 m (Table 2) indicated high specificity but low to moderate sensitivity in predicting smoke height below a certain threshold. The most important variables in the model were wind components, followed by month of observations and fire intensity within 500 km.

Several studies have used deterministic models of plume rise and

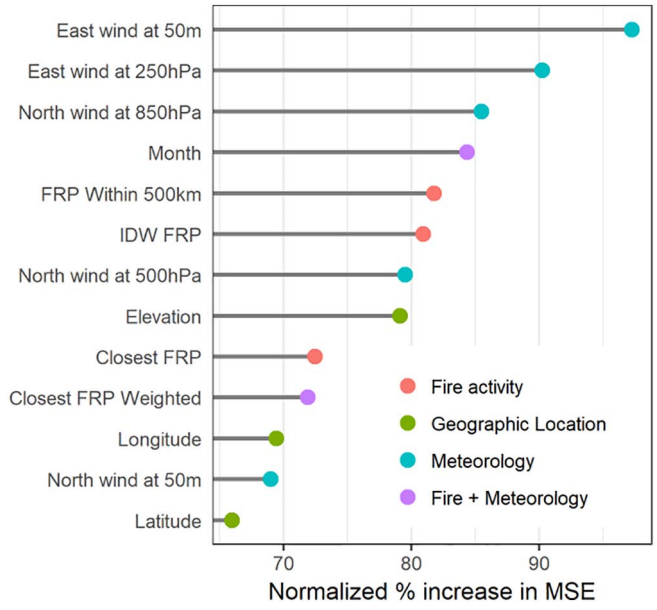

Fig. 3. Dotchart of variable importance as measured by the normalized percent increase in mean squared error (MSE) when the variable was randomly permutated. Variables are grouped into four categories related to fire activity, geographic location, meteorology and fire-meteorology combined. Only variables in the final model after the model selection procedures were included in this figure.

evaluated them against remote sensing measurements from MISR or CALIPSO (Paugam et al., 2016) but, to the best of our knowledge, ours is the first statistical model to be proposed. The R^2 value for our model was 0.82 compared with values from 0.10 to 0.30 for deterministic models of plume rise that were evaluated using observations from MISR (Raffuse et al., 2012; Sessions et al., 2011; Sofiev et al., 2012; Val Martin et al., 2012). Most of these studies predicted the initial smoke injection height over specific fires, so do not have directly comparable objectives. However, Raffuse et al. also compared the downwind plume height with CALIPSO observations, and found weak correlation for both the injection height (R^2 = 0.10) and the downwind height (R^2 = 0.22). The improved performance in our model may be due to the following factors: (1) increased predictive power due to the large dataset; (2) the application of the flexible random forests machine learning approach; and (3) the inclusion of a wide variety of fire activity, geographic location, and meteorological variables.

While the model explained much of the variability in the *Minimum Height* observations, the RMSE for the out-of-bag estimates was 560 m, which was high relative to the zone of human exposure given that the tallest building in the study area was less than 300 m. However, the results for the 100, 300, 500, and 1000 m thresholds suggest that the continuous predictions from the model also perform reasonably well when dichotomized to make binary predictions. To further evaluate this, we also built a random forests model with the same input variables to predict the binary outcome for the 300 m threshold, which resulted in more balance between the sensitivity (0.84) and specificity (0.83) when compared with those reported in here (0.65 and 0.91, respectively). This machine learning approach can be easily adapted to different types of outcomes to achieve the optimal performance depending on the purpose.

The most important variables in the model were wind patterns at different altitudes, indicating the critical influence of meteorological conditions on the vertical distribution of smoke. In this study the smoke observed by CALIPSO could have been at any stage in the dispersion process (initial emission, local dispersion, or regional transport) because the smoke-impacted data blocks were not causally associated with specific fires prior to analysis. In fact, only 73 out of the 15,617 observations were within 10 km of the closest MODIS detected fire, suggesting that most of the smoke was observed during downwind dispersion instead of initial emission. It follows that atmospheric conditions are more influential than fire intensity in our model, which is not consistent with previous studies focusing only on initial injection height (Peterson et al., 2014; Val Martin et al., 2010). Another contributor to the difference between this study and the previous studies might be that we modelled the minimum height of the smoke layer, while most previous studies modelled the average or the maximum height. The *Month* variable may be important because it reflects seasonal changes in both fire behavior and meteorological conditions.

Our dataset included much larger proportion of observations made during nighttime overpasses compared with daytime overpasses (Table 1). This seems inconsistent with previous studies that found that either (1) fire intensity and smoke emissions peaked during the day (Hodzic et al., 2007; Konovalov et al., 2013; Roberts et al., 2009), or (2) there was no obvious diurnal cycle during extreme forest fire events (Kaiser et al., 2012). However, this imbalance between nighttime and daytime smoke observations has also been reported by other studies using CALIPSO data (Huang et al., 2015), and might be attributable to differences in data quality between the nighttime and daytime products. Backscatter retrieval during daytime is substantially influenced by solar background light, resulting in sub-optimal data quality in the subsequent high-level products, such as feature classification (Hunt et al., 2009). Because we included smoke observations flagged with high data quality, a larger proportion of daytime data may have been omitted.

Although the random forests approach offers good predictive power, the interpretability of models is limited when compared with general linear regression. It is difficult to illustrate the relationship between the predictors and the outcome within the model. The partial dependence plot (Friedman, 2001) is a common way to investigate that relationship in random forests, where the marginal effect of a single predictor is shown by averaging the model predictions at each value of the predictor. Partial dependence plots (Supplementary material Fig. S1) for our model show some intuitive relationships. For example, the minimum height tends to be lower when wind speed is small, indicating more stable atmospheric conditions. However, these plots cannot depict any high-dimensional interactions between predictive variables in the model (Friedman, 2001) and any interpretation should be cautious.

We imagine the results from this model being applied in three different ways. First, they could be directly applied to existing smoke-related remote sensing products to provide another dimension of information on ground-level impacts. For example, the widely used Hazard Mapping System produces smoke plume outlines drawn by trained analysts after considering images from multiple satellites near real-time (Schroeder et al., 2008). Our predictive model could add another layer to these data and give the users a perspective on the relevance of the smoke in terms of population exposure (Fig. 4). Second, the model predictions or the variables in the model could be incorporated into operational empirical models such OSSEM, which estimate or forecast ground-level smoke concentrations using remote sensing data as the primary input. Several these models have been developed over the past few years (Price et al., 2012; Reid et al., 2015; Yao and Henderson, 2014), and better information about smoke plume height may improve their performance for public health and other applications. Finally, the predictions can be used to validate or calibrate the estimates of injection height or downwind vertical distribution from deterministic models such as BlueSky and FireWork.

There are several limitations with this study. First, the model included some meteorological variables that are readily accessible in near-real-time, but we were unable to account for the complete temporal and spatial aspects of atmospheric transport processes. However, the high predictive power, simplicity of the modelling procedure, and easy access to the variables suggest that it can be valuable for operational purposes and that it can complement conventional modelling efforts. This approach is especially useful when resources and expertise

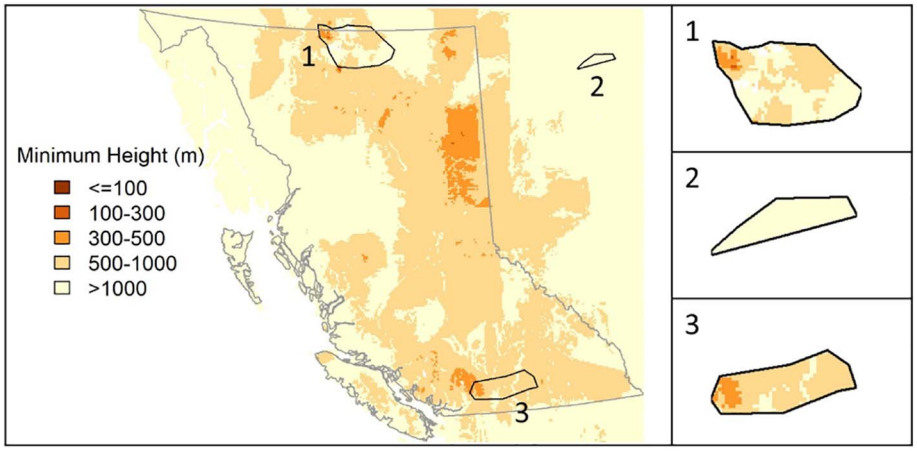

Fig. 4. Illustration of three smoke plumes from the Hazard Mapping System on July 6, 2015, showing how the minimum smoke height model can be combined with smoke plumes outlines (1, 2, and 3) to better contextualize the vertical profile. The grey line indicates the boundary of British Columbia.

are lacking to develop and maintain a complex deterministic smoke model. It can also provide valuable information from a purely data-driven perspective when our current understanding is limited or inconclusive on some physical processes of smoke emission and transport. Second, the CALIPSO vertical feature mask product has associated uncertainties and errors. Any error in its aerosol or smoke classifications would be translated into the minimum height estimates, and thus propagated into our final model. However, several studies have suggested reasonable performance of the CALIPSO classification algorithm, where the rate of correctly distinguishing aerosols and clouds is approximately 90%, with occasional misclassification of dense smoke as cloud (Amiridis et al., 2013; Liu et al., 2009). Third, we used the minimum height of any aerosol in the smoke-impacted data blocks as the minimum height of the smoke, regardless of its subtype classification. The assumption that this aerosol included smoke may not have been valid in all cases, and thus may have biased our minimum height estimates. However, this decision affected only 15% of the observations used in the model, and excluding these observations resulted in a similar model with respect to variance explained and variable importance ranking. Fourth, the training data were restricted to smoke observations with MODIS fires detected within 500 km, so the model could not account for smoke from continental transport or for smoke from small fires, smoldering fires, or fires not detected by MODIS. Although British Columbia has been affected by smoke from continental transport, a recent study reported that most smoke in the province originated from fires in the province (Brey, 2016). Small fires have always posed a challenge to models that rely on remote sensing data, but recent developments in fire detection instruments and algorithms may address this limitation in the near future (Zhang and Wooster, 2016). The limitations of MODIS fire detection could also affect the quality of the variables involving FRP in this study. Finally, by using remote sensing data to identify fires, we could not distinguish between different types of landscape fires, such as forest fires and pile burning, which might have different dynamics with respect to smoke emissions and dispersion. Hazard reduction burning of waste wood from the forest sector is common in the province, especially in the early spring and late fall, and the air quality impacts can be pronounced under some conditions (Ainslie and Jackson, 2009).

We have presented a statistical model to estimate the minimum height of forest fire smoke in the atmosphere using data on fire activity, geographic location, and meteorology. The model estimates can be used to improve the relevance of operational smoke-related remote sensing products and statistical or deterministic models intended for population exposure assessment. The machine learning approach performed well, and could easily be applied in other smoke-affected regions or for predicting other measures of the vertical distribution, such as binary indicators of smoke within the surface layer.

Acknowledgements

This work is supported by the Australian Research Council Linkage Program (LP130100146) and the British Columbia Lung Association. We also thank Dr. Martin Cope and Owen Price for their valuable feedback on the manuscript. Also, special thanks to the various divisions in the US National Aeronautics and Space Administration and Geological Survey Earth Resources Observation and Science Centre for making their data publicly available, without which this work will not be possible. We also thank the reviewers for their valuable insights that make this work better.

Appendix A. Supplementary data

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.rse.2017.12.027.

References

Ainslie, B., Jackson, P., 2009. The use of an atmospheric dispersion model to determine influence regions in the Prince George, BC airshed from the burning of open wood waste piles. J. Environ. Manag. 90, 2393–2401.
Amiridis, V., Marinou, E., Kazadzis, S., Gerasopoulos, E., Mamouri, R., Kokkalis, P., Papayannis, A., Kouremeti, N., Giannakaki, E., Liakakou, E., 2013. Evaluation of CALIPSO's Aerosol Classification Scheme during the ACEMED experimental campaign over Greece: the case study of 9th of September 2011. In: Advances in Meteorology, Climatology and Atmospheric Physics. Springer, pp. 865–871.
Atkinson, R., Kang, S., Anderson, H., Mills, I., Walton, H., 2014. Epidemiological time series studies of PM2. 5 and daily mortality and hospital admissions: a systematic review and meta-analysis. Thorax 69 (7), 660–665 (thoraxjnl-2013-204492).
B.C. Ministry of Forests, 2010. The State of British Columbia's Forests, 3rd ed. Forest Practices and Investment Branch, Victoria, BC.
Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.
Brey, S.J., 2016. Using operational hms smoke observations to gain insights on north american smoke transport and implications for air quality. In: Department of Atmospheric Science. Colorado State University, Fort Collins, Colorado.
Durán, S., Reisen, F., Rideout, K., 2014. Evidence Review: Wildfire Smoke and Public Health Risk. BC Centre for Disease Control, Vancouver BC.
EROS Data Center, 1996. GTOPO30 documentation. In: Universal Resourse Locator, Global Land Information System.
Faustini, A., Alessandrini, E.R., Pey, J., Perez, N., Samoli, E., Querol, X., Cadum, E., Perrino, C., Ostro, B., Ranzi, A., 2015. Short-term effects of particulate matter on mortality during forest fires in Southern Europe: results of the MED-PARTICLES Project. Occup. Environ. Med. 72, 323–329.
Freeborn, P.H., Wooster, M.J., Roberts, G., 2011. Addressing the spatiotemporal sampling design of MODIS to provide estimates of the fire radiative energy emitted from Africa. Remote Sens. Environ. 115, 475–489.
Freeborn, P.H., Wooster, M.J., Roy, D.P., Cochrane, M.A., 2014. Quantification of MODIS fire radiative power (FRP) measurement uncertainty for use in satellite-based active fire characterization and biomass burning estimation. Geophys. Res. Lett. 41, 1988–1994.
Freitas, S., Longo, K., Andreae, M., 2006. Impact of including the plume rise of vegetation fires in numerical simulations of associated atmospheric pollutants. Geophys. Res. Lett. 33.
Freitas, S.R., Longo, K.M., Chatfield, R., Latham, D., Silva Dias, M., Andreae, M., Prins, E., Santos, J., Gielow, R., Carvalho Jr., J., 2007. Including the sub-grid scale plume rise

of vegetation fires in low resolution atmospheric transport models. Atmos. Chem. Phys. 7, 3385–3398.
Friedman, J.H., 2001. Greedy function approximation: a gradient boosting machine. Ann. Stat. 1189–1232.
Giglio, L., 2015. MODIS Collection 6 Active Fire Product User's Guide Revision A. Department of Geographical Sciences, University of Maryland. ftp://fuoco.geog.umd.edu/modis/docs/MODIS_C6_Fire_User_Guide_A.pdf (Unpublished manuscript).
Giglio, L., Descloitres, J., Justice, C.O., Kaufman, Y.J., 2003. An enhanced contextual fire detection algorithm for MODIS. Remote Sens. Environ. 87, 273–282.
Henderson, S.B., Brauer, M., Macnab, Y.C., Kennedy, S.M., 2011. Three measures of forest fire smoke exposure and their associations with respiratory and cardiovascular health outcomes in a population-based cohort. Environ. Health Perspect. 119, 1266–1271.
Hodzic, A., Madronich, S., Bohn, B., Massie, S., Menut, L., Wiedinmyer, C., 2007. Wildfire particulate matter in Europe during summer 2003: meso-scale modeling of smoke emissions, transport and radiative effects. Atmos. Chem. Phys. 7, 4043–4064.
Huang, J., Guo, J., Wang, F., Liu, Z., Jeong, M.J., Yu, H., Zhang, Z., 2015. CALIPSO inferred most probable heights of global dust and smoke layers. J. Geophys. Res.-Atmos. 120, 5085–5100.
Hunt, W.H., Winker, D.M., Vaughan, M.A., Powell, K.A., Lucker, P.L., Weimer, C., 2009. CALIPSO lidar description and performance assessment. J. Atmos. Ocean. Technol. 26, 1214–1228.
Ichoku, C., Kaufman, Y.J., 2005. A method to derive smoke emission rates from MODIS fire radiative energy measurements. IEEE Trans. Geosci. Remote Sens. 43, 2636–2649.
Kahn, R.A., Li, W.H., Moroney, C., Diner, D.J., Martonchik, J.V., Fishbein, E., 2007. Aerosol source plume physical characteristics from space-based multiangle imaging. J. Geophys. Res.-Atmos. 112.
Kahn, R.A., Chen, Y., Nelson, D.L., Leung, F.Y., Li, Q., Diner, D.J., Logan, J.A., 2008. Wildfire smoke injection heights: two perspectives from space. Geophys. Res. Lett. 35.
Kaiser, J., Heil, A., Andreae, M., Benedetti, A., Chubarova, N., Jones, L., Morcrette, J.-J., Razinger, M., Schultz, M., Suttie, M., 2012. Biomass burning emissions estimated with a global fire assimilation system based on observed fire radiative power. Biogeosciences 9, 527.
Kollanus, V., Prank, M., Gens, A., Soares, J., Vira, J., Kukkonen, J., Sofiev, M., Salonen, R.O., Lanki, T., 2017. Mortality due to vegetation-fire originated PM 2.5 exposure in Europe–assessment for the years 2005 and 2008. Environ. Health Perspect. 125, 30–37.
Konovalov, I.B., Beekmann, M., Kaiser, J.W., Shudyaev, A.A., Yurova, A., Kuznetsova, I.N., 2013. Diurnal variations of wildfire emissions in Europe: analysis of the MODIS and SEVIRI measurements in the framework of the regional scale air pollution modelling. In: EGU General Assembly Conference Abstracts, pp. 6755.
LANCE FIRMS (2016). NASA Near Real-Time and MCD14DL MODIS Active Fire Detections. NASA (SHP format).
Larkin, N.K., O'Neill, S.M., Solomon, R., Raffuse, S., Strand, T., Sullivan, D.C., Krull, C., Rorig, M., Peterson, J., Ferguson, S.A., 2010. The BlueSky smoke modeling framework. Int. J. Wildland Fire 18, 906–920.
Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. R News 2 (3), 18–22.
Liu, Z., Omar, A., Hu, Y., Vaughan, M., Winker, D., Poole, L., Kovacs, T., 2005. CALIOP algorithm theoretical basis document. Part 3: scene classification algorithms. In: NASA-CNES Document PC-SCI-203.
Liu, Z., Vaughan, M., Winker, D., Kittaka, C., Getzewich, B., Kuehn, R., Omar, A., Powell, K., Trepte, C., Hostetler, C., 2009. The CALIPSO lidar cloud and aerosol discrimination: version 2 algorithm and initial assessment of performance. J. Atmos. Ocean. Technol. 26, 1198–1213.
Liu, J.C., Pereira, G., Uhl, S.A., Bravo, M.A., Bell, M.L., 2015. A systematic review of the physical health impacts from non-occupational exposure to wildfire smoke. Environ. Res. 136, 120–132.
Liu, J.C., Mickley, L.J., Sulprizio, M.P., Dominici, F., Yue, X., Ebisu, K., Anderson, G.B., Khan, R.F., Bravo, M.A., Bell, M.L., 2016. Particulate air pollution from wildfires in the Western US under climate change. Clim. Chang. 138, 655–666.
Lloyd, C., 2010. Spatial Data Analysis: An Introduction for GIS Users. Oxford University Press.
Mazzoni, D., Logan, J.A., Diner, D., Kahn, R., Tong, L., Li, Q., 2007. A data-mining approach to associating MISR smoke plume heights with MODIS fire measurements. Remote Sens. Environ. 107, 138–148.
Naeher, L.P., Brauer, M., Lipsett, M., Zelikoff, J.T., Simpson, C.D., Koenig, J.Q., Smith, K.R., 2007. Woodsmoke health effects: a review. Inhal. Toxicol. 19, 67–106.
Natural Resources Canada, 2016. Canadian Wildland Fire Information System: National Wildland Fire Situation Report.
Omar, A.H., Winker, D.M., Vaughan, M.A., Hu, Y., Trepte, C.R., Ferrare, R.A., Lee, K.-P., Hostetler, C.A., Kittaka, C., Rogers, R.R., 2009. The CALIPSO automated aerosol classification and lidar ratio selection algorithm. J. Atmos. Ocean. Technol. 26, 1994–2014.
Paugam, R., Wooster, M., Freitas, S., Val Martin, M., 2016. A review of approaches to estimate wildfire plume injection height within large-scale atmospheric chemical transport models. Atmos. Chem. Phys. 16, 907–925.
Pavlovic, R., Chen, J., Anderson, K., Moran, M.D., Beaulieu, P.-A., Davignon, D., Cousineau, S., 2016. The FireWork air quality forecast system with near-real-time biomass burning emissions: recent developments and evaluation of performance for the 2015 North American wildfire season. J. Air Waste Manage. Assoc. 66, 819–841.
Peterson, D., Hyer, E., Wang, J., 2014. Quantifying the potential for high-altitude smoke injection in the North American boreal forest using the standard MODIS fire products and subpixel-based methods. J. Geophys. Res.-Atmos. 119, 3401–3419.
Pope III, C.A., Dockery, D.W., 2006. Health effects of fine particulate air pollution: lines that connect. J. Air Waste Manage. Assoc. 56, 709–742.
Price, O.F., Williamson, G.J., Henderson, S.B., Johnston, F., Bowman, D.M., 2012. The relationship between particulate pollution levels in Australian cities, meteorology, and landscape fire activity detected from MODIS hotspots. PLoS One 7, e47327.
Raffuse, S.M., Craig, K.J., Larkin, N.K., Strand, T.T., Sullivan, D.C., Wheeler, N.J., Solomon, R., 2012. An evaluation of modeled plume injection height with satellite-derived observed plume height. Atmosphere 3, 103–123.
Reid, C.E., Jerrett, M., Petersen, M.L., Pfister, G.G., Morefield, P.E., Tager, I.B., Raffuse, S.M., Balmes, J.R., 2015. Spatiotemporal prediction of fine particulate matter during the 2008 northern California wildfires using machine learning. Environ. Sci. Technol. 49, 3887–3896.
Reid, C.E., Brauer, M., Johnston, F.H., Jerrett, M., Balmes, J.R., Elliott, C.T., 2016a. Critical review of health impacts of wildfire smoke exposure. Environ. Health Perspect. 124, 1334–1343.
Reid, C.E., Jerrett, M., Tager, I.B., Petersen, M.L., Mann, J.K., Balmes, J.R., 2016b. Differential respiratory health effects from the 2008 northern California wildfires: a spatiotemporal approach. Environ. Res. 150, 227–235.
Rienecker, M.M., Suarez, M.J., Gelaro, R., Todling, R., Bacmeister, J., Liu, E., Bosilovich, M.G., Schubert, S.D., Takacs, L., Kim, G.-K., 2011. MERRA: NASA's modern-era retrospective analysis for research and applications. J. Clim. 24, 3624–3648.
Roberts, G., Wooster, M., Lagoudakis, E., 2009. Annual and diurnal African biomass burning temporal dynamics. Biogeosciences 6.
Sakiyama, S., 2013. The Bluesky Western Canada Wildfire Smoke Forecasting System. British Columbia Ministry of Environment, Victoria, BC.
Schroeder, W., Ruminski, M., Csiszar, I., Giglio, L., Prins, E., Schmidt, C., Morisette, J., 2008. Validation analyses of an operational fire monitoring product: the Hazard Mapping System. Int. J. Remote Sens. 29, 6059–6066.
Sessions, W., Fuelberg, H., Kahn, R., Winker, D., 2011. An investigation of methods for injecting emissions from boreal wildfires using WRF-Chem during ARCTAS. Atmos. Chem. Phys. 11, 5719–5744.
Shepard, D., 1968. A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM National Conference. ACM, pp. 517–524.
Sofiev, M., Ermakova, T., Vankevich, R., 2012. Evaluation of the smoke-injection height from wild-land fires using remote-sensing data. Atmos. Chem. Phys. 12, 1995–2006.
Spracklen, D.V., Mickley, L.J., Logan, J.A., Hudman, R.C., Yevich, R., Flannigan, M.D., Westerling, A.L., 2009. Impacts of climate change from 2000 to 2050 on wildfire activity and carbonaceous aerosol concentrations in the western United States. J. Geophys. Res.-Atmos. 114.
Stull, R.B., 2012. An Introduction to Boundary Layer Meteorology. Springer Science & Business Media.
US NASA, 2016. Data User's Guide - Data Product Descriptions - Lidar Level 25 km Vertical Feature Mask (VFM) Product.
Val Martin, M., Logan, J.A., Kahn, R.A., Leung, F.-Y., Nelson, D.L., Diner, D.J., 2010. Smoke injection heights from fires in North America: analysis of 5 years of satellite observations. Atmos. Chem. Phys. 10, 1491–1510.
Val Martin, M., Kahn, R.A., Logan, J.A., Paugam, R., Wooster, M., Ichoku, C., 2012. Space-based observational constraints for 1-D fire smoke plume-rise models. J. Geophys. Res.-Atmos. 117.
van Etten, J., 2012. gdistance: distances and routes on geographical grids. URL. http://CRAN.R-project.org/package=gdistance (R package version, 1.1-4).
Vaughan, M., Winker, D.M., Powell, K., 2005. CALIOP algorithm theoretical basis document, part 2: feature detection and layer properties algorithms. Rep. PC-SCI 202, 87.
Vaughan, M.A., Powell, K.A., Winker, D.M., Hostetler, C.A., Kuehn, R.E., Hunt, W.H., Getzewich, B.J., Young, S.A., Liu, Z., McGill, M.J., 2009. Fully automated detection of cloud and aerosol layers in the CALIPSO lidar measurements. J. Atmos. Ocean. Technol. 26, 2034–2050.
Winker, D.M., Hostetler, C., Vaughan, M., Omar, A., 2006. CALIOP Algorithm Theoretical Basis Document, Part 1: CALIOP Instrument, and Algorithms Overview. Release, 2, 29.
Wooster, M., Zhukov, B., Oertel, D., 2003. Fire radiative energy for quantitative study of biomass burning: derivation from the BIRD experimental satellite and comparison to MODIS fire products. Remote Sens. Environ. 86, 83–107.
Wooster, M.J., Roberts, G., Perry, G., Kaufman, Y., 2005. Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. J. Geophys. Res.-Atmos. 110.
Yao, J., Henderson, S.B., 2014. An empirical model to estimate daily forest fire smoke exposure over a large geographic area using air quality, meteorological, and remote sensing data. J. Expo. Sci. Environ. Epidemiol. 24, 328.
Yao, J., Eyamie, J., Henderson, S.B., 2016. Evaluation of a spatially resolved forest fire smoke model for population-based epidemiologic exposure assessment. J. Expo. Sci. Environ. Epidemiol. 26, 233–240.
Zhang, T., Wooster, M., 2016. Small Fire Detection Algorithm development using VIIRS 375m Imagery: application to agricultural fires in eastern China. In: EGU General Assembly Conference Abstracts, pp. 17722.